Statistica Sinica 2(1992), 47-68

ON ASYMPTOTIC EFFICIENCY IN ESTIMATION THEORY

Wing Hung Wong

The University of Chicago

Abstract: This is an account of the mathematical formulation of asymptotic efficiency in estimation theory from the point of view of the concentration of the estimators around the true parameter value. The purpose is not to propose any new definition of efficiency, but rather to consider the inter-relationships among some rather scattered existing results with the aim of connecting them into a coherent whole. In the process of doing so, some improvements and extensions will also be given. The theory is developed first in the case of a scalar parameter. It is then extended by a simple argument to cover the estimation of a scalar function of a multi or infinite dimensional parameter.

Key words and phrases: Fisher's information bound, concentration probability, regular estimators, locally asymptotically minimax.

1. Introduction

After the pioneering works of Edgeworth (1908-9) and Fisher (1925), it was believed that the maximum likelihood estimator (MLE) $\hat{\theta}_n$ is asymptotically optimal among a large class of estimators. According to Pratt (1976), both Edgeworth and Fisher gave more or less correct proofs of the optimality of MLE (for a location parameter) among the class of M-estimators, and each of them also recognized that the optimality actually extends to more general cases. The general result that Fisher had in mind (and attempted to prove) may be formulated mathematically as follows:

(i) $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} N(0, i_\theta^{-1})$ under $P_\theta(\mathbf{X}_n)$.

(ii) If T_n is any asymptotically normal estimator (i.e. $\sqrt{n}(T_n - \theta) \xrightarrow{\mathcal{D}} N(0, v_\theta)$ under $P_\theta(\mathbf{X}_n)$), then $v_\theta \geq i_\theta^{-1}$.

Here $P_\theta(\mathbf{X}_n)$ denotes the distribution of the data $\mathbf{X}_n = (X_1, \ldots, X_n)$ where $X_1, X_2, \ldots$ are independently identically distributed (i.i.d) in R according to a density $f_\theta(\cdot)$ indexed by a real valued parameter θ, and i_θ is the Fisher information of the family $\{f_\theta(\cdot)\}$, defined more precisely below. Part (i) of Fisher's claim is typically true under mild conditions on the family of densities $\{f_\theta(\cdot)\}$.

However, an example by J. L. Hodges, Jr. showed that part (ii) of Fisher's claim is not true as stated even in the smoothest problems. In Hodge's example, an asymptotically normal estimator T_n is constructed whose asymptotic variance v_θ has the property that $v_\theta \leq i_\theta^{-1}$ for all θ, and $v_{\theta_0} < i_{\theta_0}^{-1}$ for a particular value θ_0. Hodge's example was reported in LeCam (1953), who called such an estimator "superefficient" at θ_0. He then proceeded to establish, among other things, the celebrated result that the set of θ where superefficiency occurs is of measure zero.

To state the results in more precise terms, we need to introduce a few notations and regularity conditions. For simplicity, in Sections 1 to 5 we assume that the parameter space is the unit interval, i.e. $\theta \in \Omega = [0,1]$. Let $\mathbf{X}_n = (X_1, \ldots, X_n)$ be the observation vector taking values in a sample space $(\mathcal{X}_n, \mathcal{B}_n, \nu_n)$ where ν_n is a σ-finite measure on the σ-field $\mathcal{B}_n$. Denote by $P_\theta(\mathbf{X}_n)$ the distribution of $\mathbf{X}_n$ under θ, and $p_\theta(\mathbf{x}_n)$ the corresponding density function with respect to ν_n. An estimator of θ based on $\mathbf{X}_n$ is by definition simply a measurable function $T_n = T_n(\mathbf{X}_n)$. In this paper, we study the properties of estimators T_n as the "sample size" n tends to infinity, under one or more of the following conditions on the densities $p_\theta(\mathbf{x}_n)$.

Condition (M). *For every n, $p_\theta(\mathbf{x}_n)$ is a measurable function in $(\theta, \mathbf{x}_n)$.*

Condition (L). *For all $h \in R$, $\frac{dP_{\theta+h/\sqrt{n}}}{dP_\theta}(\mathbf{X}_n) = e^{h\Delta_{n,\theta} - \frac{1}{2} i_\theta h^2 + R_n(\theta,h)}$ where, under $P_\theta(\mathbf{X}_n)$, $\Delta_{n,\theta} \xrightarrow{\mathcal{D}} N(0, i_\theta), R_n(\theta, h) \xrightarrow{\mathcal{P}} 0$; and i_θ is a strictly positive and continuous function of θ.*

Condition (P). *For any piecewise continuous prior density $\pi(\cdot)$ on Ω, let $I_\pi =$ interior of $\{\theta : \pi(\theta) > 0\}$; then, for $\theta \in I_\pi$ and any $s \in R$,*

$$P(\sqrt{n}(\Theta - \hat{\theta}_n) < s \mid \mathbf{X}_n; \pi) \xrightarrow{\mathcal{P}} P(Z_\theta < s) \text{ under } P_\theta(\mathbf{X}_n).$$

Here $\hat{\theta}_n$ denotes a maximum likelihood estimator of θ based on $\mathbf{X}_n$, Θ denotes the random variable whose joint density with $\mathbf{X}_n$ is given by $p(\theta, \mathbf{x}_n) = \pi(\theta) p_\theta(\mathbf{x}_n)$, Z_θ denotes a $N(0, i_\theta^{-1})$ variable, and i_θ is a strictly positive and continuous function in θ.

Condition (L) is LeCam's "locally asymptotically normal" (LAN) condition (see LeCam (1960)). Condition (P), in essence, requires the posterior distribution of θ to be asymptotically normal when the data is generated from a parameter value within the support of the prior. In the case when $X_1, \ldots, X_n$ are i.i.d. with a common density $f_\theta(x)$, we can usually take $\Delta_{n,\theta}$ to be the normalized score function and i_θ to be the Fisher information of $\{f_\theta(\cdot)\}$, i,e.

$$\Delta_{n,\theta} = n^{-1/2} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f_\theta(X_i), \quad i_\theta = E_\theta\Big(\frac{\partial}{\partial\theta} \log f_\theta(X)\Big)^2;$$

then these conditions are typically satisfied under mild conditions on the family $\{f_\theta(\cdot)\}$. In this paper, however the derivation of conditions (L) and (P) from more elementary conditions will not be discussed further. It seems that conditions (M), (L) and (P) are the natural conditions to use in more complex situations, such as when the observations are dependent; whereas the simpler conditions in the i.i.d. case that imply these conditions are more difficult to generalize. Since all the difficulties in formulating a satisfactory concept of asymptotic efficiency are present even in the smoothest problems, for example, when $X_1, X_2, \ldots$ are i.i.d. $N(\theta, 1)$, it is felt that the imposition of the above conditions on the parametric family does not distract us from the central issues.

We can now state, in mathematical terms, the aforementioned result that superefficiency can occur only in sets of measure zero. LeCam (1953) established such a result for asymptotically efficient estimators when efficiency is defined in terms of a risk. That this result holds in fact for asymptotically normal estimators with asymptotic variances as the criterion was first established by Bahadur (1964).

Proposition 1. *Under conditions* (M) *and* (L), *if* $\sqrt{n}(T_n - \theta) \xrightarrow{\mathcal{D}} N(0, v_\theta)$ *for all* $\theta \in \Omega$, *then* $\mu\{\theta : v_\theta < i_\theta^{-1}\} = 0$, *where* μ *denotes the Lebesgue measure on* Ω.

This result has been extended in several ways. First of all, the restriction of the comparisons to only the class of asymptotically normal estimators seems too restrictive. But how can one compare estimators having different types of asymptotic distributions? A natural way, studied previously by several authors (for example, Basu (1956)), is to compare their concentration probabilities around the true parameter value. Specifically, since the MLE or other supposedly asymptotically efficient estimators have $N(0, i_\theta^{-1})$ as their asymptotic distribution, we may want to establish, for a class of estimator $\{T_n\}$, that the concentration probability $P_\theta(\sqrt{n}|T_n - \theta| < \rho)$ is asymptotically bounded above by $P(|Z_\theta| < \rho)$. The sense in which the latter can be interpreted as an asymptotic upper bound for the former depends on how large is the class of estimators under consideration. The larger this class of estimators, the weaker will be the resulting "Fisher's bound" for the concentration probabilities. This is a rather trivial point, yet it serves as a useful theme in the organization of the results reviewed in this paper. In Section 2, we fix a parameter value θ_0 and study conditions on T_n which are stringent enough that Fisher's bound holds at θ_0. The essence of the regularity condition is that some aspect of the distribution of $\sqrt{n}(T_n - \theta_n)$ under any sequence $\{\theta_n\}$ of "local alternatives" of θ_0 must converge to a limit independent of the sequence $\{\theta_n\}$. An estimator satisfying such a condition will be said to be "regular" at θ_0. Thus, regularity at θ_0 excludes superefficiency at θ_0.

In Section 3, we study the larger class of "stable estimators", i.e. estimators having an asymptotic distribution at each θ. Under stability it is shown that Fisher's bound for the asymptotic distribution holds for almost all θ. This is a natural generalization of Proposition 1 above.

In Section 4, we remove all conditions on T_n, i.e. the class of estimators is now the largest possible. As a result, the sense of Fisher's bound must be further weakened. In general, the most one can conclude is that Fisher's bound is satisfied in some "locally averaged" sense. The relationship between this and Hajek's "locally asymptotically minimax" (LAM) form of Fisher's bound will be examined. We also study, for large n, the "size" of the set of θ value where Fisher's bound is exceeded by an amount $\epsilon > 0$. The material in Sections 2 and 3 is not essential for understanding Section 4.

In Sections 5 and 6, we present an extension of Fisher's bound for concentration probabilities of regular estimators of a scalar functional of a possibly infinite dimensional parameter.

Although only a small portion of the results are new, proofs are provided for all propositions stated in this paper. These proofs are collected in Section 7 so as not to disrupt the main exposition. It is the author's hope that this will render the account self-contained.

2. Regular Estimators

In this section we fix a particular value θ_0 in the interior of Θ and study conditions on T_n that guarantee that Fisher's bound is satisfied at the value θ_0. Hodges's example showed that it is not enough to require only that $\sqrt{n}(T_n - \theta_0)$ has an asymptotic distribution; something more is needed. The key observation was made in Bahadur (1964). We can reformulate his result, given only for asymptotically normal estimators, in terms of concentration probabilities.

Proposition 2 (Bahadur). *Under condition* (L), *if T_n is asymptotically median unbiased under local alternatives, i.e. for all* (*sequences of*) *local alternatives of the form* $\theta_{n,h} = \theta_0 + h/\sqrt{n}$,

$$P_{\theta_{n,h}}(T_n < \theta_{n,h}) \to \frac{1}{2},$$

then for all $a > 0$, $b > 0$,

$$\limsup_{n\to\infty} P_{\theta_0}\Big(- a < \sqrt{n}(T_n - \theta_0) < b\Big) \leq P(-a < Z_{\theta_0} < b).$$

This is the first result of this type, namely, if the distribution of T_n is in some sense regular as $n \to \infty$ under local alternatives of a point θ_0, then Fisher's bound holds at that θ_0. It will be established below that the median unbiasedness condition can be relaxed substantially. However, let us first compare Proposition 2 to another famous result, namely, the "representation theorem" for asymptotically "regular" estimators, due to Hajek (1970).

Proposition 3 (Hajek). *Under condition* (L), *if*

(i) $\sqrt{n}(T_n - \theta_{n,h})$ *has a limiting distribution* $F_h(\cdot)$ *under any local alternative* $\{\theta_{n,h} = \theta_0 + h/\sqrt{n}\}$, *i.e., for any* $h \in R$,

$$P_{\theta_{n,h}}\left(\sqrt{n}(T_n - \theta_{n,h}) < s\right) \to F_h(s) \quad \forall\ s\ \in R$$

and

(ii) $F_h(\cdot)$ *is independent of h, with common limiting distribution* $F^*(\cdot)$, *say, then* F^* *is the convolution of a* $N(0, i_{\theta_0}^{-1})$ *distribution with some other distribution function G.*

The condition in Proposition 3 requires $P_{\theta_{n,h}}\left(T_n - \theta_{n,h}\ < \frac{s}{\sqrt{n}}\right)$ to converge to a limit independent of h for each s, whereas Proposition 2 needs this only for $s = 0$. In this sense Proposition 3 requires a much stronger condition than Proposition 2, but as a consequence it yields a characterization of the whole asymptotic distribution as opposed to only bounds for concentration probabilities. Moreover, the limit of $P_{\theta_{n,h}}\left(T_n\ < \theta_{n,h}\right)$ is required to be 1/2 in Proposition 2, whereas in Proposition 3 it is allowed to be any constant independent of h. This suggests that Fisher's bound for probabilities of concentration may still be valid under a weaker condition than both.

Proposition 4. *Under condition* (L), *if for all rational* $\rho > 0$,

$$\limsup_{n\to\infty} P_{\theta_{n,\rho}}(T_n < \theta_{n,\rho}) \le \liminf_{n\to\infty} P_{\theta_{n,-\rho}}(T_n < \theta_{n,-\rho}),$$

then, for any $\rho > 0$,

$$\limsup P_{\theta_0}(\sqrt{n}|T_n - \theta_0| < \rho) \le P(|Z_{\theta_0}| < \rho).$$

An estimator satisfying the condition of Proposition 4 is said to be *regular* at θ_0. Note that the condition in Proposition 4 is weaker than those of both Propositions 2 and 3. Of course, the conclusion in Proposition 4 is weaker than both also. In particular, while the lower bound for probabilities of concentration in Proposition 2 applies to any interval containing zero, the bound in Proposition 4 is applicable only to symmetric intervals. However, the example below shows that, in the absence of asymptotic median unbiasedness, the probabilities of concentration (of $\sqrt{n}(T_n - \theta)$) in asymmetric intervals are really not relevant for the concept of efficiency.

Example 1. Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(\theta, 1)$, (then $i_\theta \equiv 1$). Consider the estimator $T_n = \overline{X}_n + \frac{a}{\sqrt{n}}$ where $a > 0$ is a constant. Clearly, T_n is an inferior estimate to the sample mean $\overline{X}_n$. However, for all $\epsilon > 0$ sufficiently small, the interval $I = (-\epsilon, 2a - \epsilon)$ has the property that $P_\theta(\sqrt{n}(T_n - \theta) \in I) > P_\theta(\sqrt{n}(\overline{X}_n - \theta) \in I)$ for all $\theta \in \Omega$ and for all n.

Remark. While we are on the subject of "regular estimates", it may be worthwhile to comment that the proof of Proposition 3, presented in Section 5 below, provides a scheme whereby other theorems can be deduced for the asymptotic distributions of estimators satisfying condition (i) of Proposition 3. As an example, consider

Proposition 5. *Under condition* (L), *suppose*

(i) $\sqrt{n}(T_n - \theta_{n,h})$ *has a limiting distribution* $F_h(\cdot)$ *under any local alternative* $\{\theta_{n,h} = \theta_0 + h/\sqrt{n}\}$,

(ii) *The distributions* $F_h(\cdot)$ *have a common mean value* a (*i.e.* $a \equiv \int y dF_h(y)$ *exists and is independent of* h).

Denote by Y_0 *a random variable with the same distribution as the limiting distribution of* $\sqrt{n}(T_n - \theta_0)$ *under* θ_0*; then* $E(Y_0^2) \geq a^2 + i_{\theta_0}^{-1}$.

These are the weakest conditions known to the author under which Fisher's bound for asymptotic variances is guaranteed to hold at an arbitrary value θ_0.

3. Stable Estimators

An (sequence of) estimator T_n is said to be *stable* if for all θ, $\sqrt{n}(T_n - \theta)$ has a limiting distribution under θ. Clearly, a stable estimator need not be regular at all θ.

Proposition 1 suggests that for stable estimators (which includes asymptotically normal estimators), Fisher's bound pertaining to the asymptotic distribution, though not necessarily valid for all θ, might be valid for almost all θ. There are two general ways of obtaining such a result. One approach, introduced in Bahadur (1964), is to first establish regularity conditions that guarantee Fisher's bound at any given point, and then show that stability implies that these regularity conditions are satisfied (along a subsequence) almost everywhere. Using this argument and Proposition 2, he was able to give a short proof of Proposition 1. Using the same approach and Proposition 4, we obtain the following extension of Proposition 1.

Proposition 6. *Under conditions* (M) *and* (L), *if* T_n *is stable and if* $\gamma(\theta) = \lim P_\theta(T_n < \theta)$ *is continuous almost everywhere, then, for almost all* θ *in* Ω,

$$\lim P_\theta(\sqrt{n}|T_n - \theta| < \rho) \leq P(|Z_\theta| < \rho) \quad \textit{for all } \rho \geq 0.$$

Another approach for obtaining the almost everywhere Fisher bound for stable estimators is to first establish a weaker version of Fisher's bound, such as the "locally averaged" version in Section 4 below, and then argue that stability allows one to improve this bound to an "almost everywhere" version. This idea appears to have been introduced first by LeCam (1953). Making use of Proposition 12 below, this approach leads to Proposition 7.

Proposition 7. *Under conditions* (M) *and* (P), *if* T_n *is a stable estimator, then for almost all* $\theta \in \Omega$,

$$\lim P_\theta(\sqrt{n}|T_n - \theta| < \rho) \leq P(|Z_\theta| < \rho) \quad \textit{for all } \rho \geq 0.$$

It appears that stability is close to the weakest regularity condition on T_n for Fisher's bound to hold for almost all θ. For any statistical experiment, if the distribution of T_n is allowed to oscillate, Fisher's bound may in fact be violated everywhere!

Example 2*. For any $\rho > 0$, define a sequence of numbers $\{a_n\}$ as follows: $a_0 = 0$,

$$a_n = \left(a_{n-1} + \frac{\rho}{2\sqrt{n}}\right) - \left(\text{the integer part of } a_{n-1} + \frac{\rho}{2\sqrt{n}}\right).$$

In other words, (for large enough n) a_n is obtained by adding $\rho/2\sqrt{n}$ to a_{n-1}, but "wrapped around" if the resulting sum exceeds 1. Since $\Sigma_1^\infty \frac{1}{\sqrt{n}} = \infty$, the sequence $\{a_n\}$ will transverse the interval $\Omega = [0, 1]$ an infinite number of times. Thus, any $\theta \in \Omega$ will be covered by an interval of the form $[a_{n-1}, a_n)$ infinitely often. Now define the estimator T_n by choosing a point randomly from $[a_{n-1}, a_n)$. This is clearly a ridiculous estimator. Yet, irrespective of the form of $\{p_\theta(\cdot)\}$, Fisher's bound is violated for all $\theta \in \Omega$, because

$$\limsup P_\theta(\sqrt{n}|T_n - \theta| < \rho) = 1 > P(|Z_\theta| < \rho) \quad \text{for all } \theta \in \Omega.$$

4. Arbitrary Estimators

*The author was informed by Professor L. LeCam that he and his coworkers had already obtained a similar example of this phenomenon, which is reported in a Berkeley thesis of N. H. Cheng.

Although the conditions in Propositions 6 and 7 are sufficiently mild to cover most estimators that we may encounter in practice, from a logical point of view, we may still wish to have a concept of asymptotic efficiency covering all estimators. For simplicity of notation, let $\rho > 0$ be arbitrary but fixed, and let

$$g_n(\theta) = P_\theta(\sqrt{n}|T_n - \theta| < \rho)$$

and

$$g(\theta) = P(|Z_\theta| < \rho).$$

Example 2 above shows that for an arbitrary T_n, there is no hope of trying to bound $\limsup g_n(\theta)$ by $g(\theta)$. Hence it is necessary to further weaken the sense of Fisher's bound. One way to formulate a weaker version of Fisher's bound is to establish that, for any $\epsilon > 0$ and for all n sufficiently large, the sets

$$A_n = \{\theta : g_n(\theta) \leq g(\theta) + \epsilon\}$$

are large subsets of Ω.

First of all, we must be sure that regular or stable estimators would automatically satisfy such a weaker form of Fisher's bound.

Proposition 8. *Under conditions* (M) *and* (P), *if T_n is a stable estimator, then* $\mu(A_n) \to 1$.

What can be said about the size of A_n when T_n is an arbitrary estimator? An important result which bears on this question was given by Hajek (1972).

Proposition 9 (Hajek). *Suppose condition* (P) *holds, then for any $\theta_0 \in \Omega$,*

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \Big(\inf_{|\theta - \theta_0| \leq \delta} g_n(\theta) \Big) \leq g(\theta_0).$$

Hajek (1972) actually proved this result under condition (L) but not (P). We stated it under (P) so that it relates better with the other results in this section. Hajek (1972) also contained references to earlier contributors to the development of this 'locally asymptotically minimax' (LAM) form of Fisher's bound, including L. LeCam, C. Stein, H. Rubin and H. Chernoff. Proposition 9 is equivalent to the following proposition.

Proposition 10. *Suppose condition* (P) *holds; then for any $\epsilon > 0$, the set $A_n = \{\theta : g_n(\theta) \leq g(\theta) + \epsilon\}$ becomes dense in Ω as $n \to \infty$ (i.e. for any interval $I \subset \Omega, A_n \cap I \neq \phi$ for all sufficiently large n.)*

Thus, the LAM form of Fisher's bound, which is applicable to arbitrary estimators, is considerably weaker than the form available for stable estimators.

In fact, it does not even exclude the possibility that $\mu(A_n) = 0$, $\forall n$. Can we say more about the size of A_n for arbitrary estimators? The answer is yes, but just by a bit more.

Proposition 11. *Suppose condition* (P) *holds; then for any interval* $I \subset \Omega$,

$$\liminf_{n\to\infty} \mu(A_n \cap I) > 0.$$

Note that Proposition 10 still does not say that A_n is large in Lebesgue measure if n is large. It is tempting for us to go further and attempt to establish that $\mu(A_n) \to 1$ for arbitrary T_n. The following example shows that such an effort will be futile. The idea of trying a purturbation by a scaled sinusoid in this example was suggested by Professor Charles Stein when the author posed the problem to him in a private conversation. To explain the example, we need a technical lemma.

Lemma. *Let* $Z \sim N(0,1)$; *then for all* ϵ, ρ *and* $|\gamma|$ *sufficiently small, say,* $0 < \epsilon < \epsilon_0, 0 < \rho < \rho_0, 0 \le |\gamma| < \gamma_0$, *we have*

$$P\left(|Z - \frac{1}{2}\sin(Z+\gamma)| < \rho\right) > P(|Z| < \rho) + \epsilon.$$

Example 3. Let $X_1, X_2, \ldots$ be i.i.d. $N(\theta, 1), \theta \in \Omega = [0,1]$, $\overline{X}_n = n^{-1}\sum_1^n X_i$,

$$\begin{aligned} T_n &= \overline{X}_n - \frac{1}{2\sqrt{n}}\sin(\sqrt{n}\,\overline{X}_n), \\ g_n(\theta) &= P_\theta(\sqrt{n}|T_n - \theta| < \rho) \\ g(\theta) &= P(|Z| < \rho) \\ B_n &= \{\theta \in \Omega :\ g_n(\theta) > g(\theta) + \epsilon\}. \end{aligned}$$

Suppose $\epsilon < \epsilon_0$, $\rho < \rho_0$ where ϵ_0, ρ_0, γ_0 are as in the above lemma; then $\mu(B_n) > \gamma_0/2\pi$ for all sufficiently large n. Hence, in this example, with ϵ and ρ chosen as above, the sets B_n where Fisher's bound fails have Lebesgue measures bounded away from zero. To see this, write

$$\sqrt{n}(T_n - \theta) = \sqrt{n}(\overline{X}_n - \theta) - \frac{1}{2}\sin\left(\sqrt{n}(\overline{X}_n - \theta) + \sqrt{n}\theta\right).$$

Thus, $\sqrt{n}(T_n - \theta)$ has the same distribution as

$$Z - \frac{1}{2}\sin(Z + \sqrt{n}\theta) \quad \text{where} \quad Z \sim N(0,1).$$

Suppose θ satisfies

$$(*) \qquad |\sqrt{n}\theta - 2\pi k| < \gamma_0 \quad \text{for some integer} \quad k \in \left(0, \frac{\sqrt{n}}{2\pi}\right).$$

Then, by the lemma, we must have

$$P_\theta(\sqrt{n}|T_n - \theta| < \rho) > P(|Z| < \rho) + \epsilon.$$

Hence, B_n contains all $\theta \in [0,1]$ satisfying $(*)$, i.e. all $\theta \in [0,1]$ satisfying $|\theta - 2\pi\frac{k}{\sqrt{n}}| < \gamma_o/\sqrt{n}$, $k = 1,2,\ldots[\frac{\sqrt{n}}{2\pi}]$, where $[\frac{\sqrt{n}}{2\pi}]$ is the largest integer below $\sqrt{n}/2\pi$. Therefore, $\mu(B_n) \geq ([\sqrt{n}/2\pi])(2\gamma_0/\sqrt{n}) \to \gamma_0/\pi$ as $n \to \infty$.

In the above discussion we saw that the set $A_n = \{\theta : g_n(\theta) < g(\theta) + \epsilon\}$ where Fisher's bound holds becomes dense and has non-ignorable measure when n is large. However, we found that, in general, A_n need not have large Lebesgue measure. Is it possible to obtain a stronger form of Fisher's bound while still imposing no condition on T_n?

Recall that our objective is to establish, for any $\epsilon > 0$, that the function $f_n(\theta) = g_n(\theta) - g(\theta)$ is bounded above by ϵ when n is sufficiently large. It is seen in Example 2 and 3 that the difficulty in obtaining such a result lies in the fact that $f_n(\cdot)$ can be highly irregular (non-smooth) function. Now, an important lesson in modern mathematics is this: when it is difficulty to study a sequence of irregular functions directly, study smoothed versions of them first. Thus, although we cannot bound $f_n(\cdot)$ directly, we might be able to bound a smoothed version of it as n becomes large. The following result shows that this is indeed the case, no matter how small the degree of smoothing. In fact, it is this result that leads to the most straightforward proofs of all proceeding propositions in this section.

Proposition 12. *Suppose condition* (P) *holds; then for any* θ_0 *in the interior of* Ω, *and any* $\epsilon > 0, \delta > 0$, *it is true that*

$$\frac{1}{2\delta}\int_{|\theta-\theta_0|<\delta} (g_n(\theta) - g(\theta))\,d\theta < \epsilon$$

for all sufficiently large n.

It also follows easily from this result that

Proposition 13. *Under condition* (P), *if* θ_0 *is in the interior of* Ω, *then*

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\left(\frac{1}{2\delta}\int_{|\theta-\theta_0|<\delta} g_n(\theta)\,d\theta\right) \leq g(\theta_0).$$

Proposition 13 is analogous to Hajek's LAM result (Proposition 9), except that local infinmum is replaced by local averaging, rendering a stronger sense for Fisher's bound.

Some versions of Proposition 12 and 13 were probably known very early previously (see especially LeCam (1953, 1958)). However, these results appear not to have been appreciated as much as they deserve to be, as is evident from the dominance of the LAM formulation in the current literature.

In the author's opinion, the strongest possible form of Fisher's bound for arbitrary estimators may in fact be formulated using local averaging (of the concentration probability $g_n(\cdot)$). However, Proposition 12 and 13 can still be improved in various ways. For example, it is more useful to have the bound in Proposition 12 uniform over θ_0. Also, for the results to be relevant, practically, we need to have some idea of how large the sample size needs to be for the bound to be valid. To obtain such results, stronger conditions than condition (P) may be required for the parametric family of distributions, though no extra condition on the estimator T_n is needed.

5. Estimation of Differentiable Functions

Let $\lambda(\theta)$ be a real valued, continuously differentiable function of θ, i.e.

$$\lim_{\theta_1 \to \theta} \frac{\lambda(\theta_1) - \lambda(\theta)}{\theta_1 - \theta} = \lambda'(\theta) \quad \text{exists and is continuous in } \theta.$$

To establish bounds for estimating such a function, the family of distributions $\{P_\theta(\mathbf{X}_n)\}$ needs to satisfy a condition slightly stronger than condition (L).

Condition (L′). Condition (L) is satisfied with $\theta + h/\sqrt{n}$ replaced by $\theta_{n,h} = \theta + t_n h/\sqrt{n}$, for any $t_n \to 1$.

Proposition 14. *Suppose condition* (L′) *is satisfied at* $\theta = \theta_0$ *and* $\lambda(\theta)$ *is a continuously differentiable function in a neighborhood of* θ_0*, with derivative* $\lambda'(\theta_0) \neq 0$. *Let* T_n *be a regular estimator of* $\lambda(\theta)$ *in the sense of Proposition* 4, *i.e. for all* $\rho > 0$

$$\limsup P_{\theta_{n,\rho}}(T_n < \lambda(\theta_{n,\rho})) \leq \liminf P_{\theta_{n,-\rho}}(T_n < \lambda(\theta_{n,-\rho}))$$

where $\theta_{n,\rho}$ *is as defined in condition* (L′). *Then*

$$\limsup P_{\theta_0}(\sqrt{n}|T_n - \lambda(\theta_0)| < \rho) \leq P(|Z_0| < \rho)$$

where

$$Z_0 \sim N(0,\ \lambda'(\theta_0)^2 \cdot i_{\theta_0}^{-1}).$$

6. Infinite Dimensional Parameter Space

In this section it is assumed that the family of probabilities $\{P_\phi\}$ is indexed by a possibly infinite dimensional parameter $\phi \in \Phi$ where the parameter space Φ is a subset of a linear space $\mathcal{L}$. Denote the true parameter value by ϕ_0.

Let $V_0 \subset \mathcal{L}$ be the set of $\mathbf{v} \in \mathcal{L}, \mathbf{v} \neq 0$ such that

i) there exists an $\epsilon > 0$ such that

$$\phi_0 + t\mathbf{v} \in \Phi \quad \forall\ |t| < \epsilon,$$

ii) the one dimensional family $\{P_{\phi_0+t\mathbf{v}},\ |t| < \epsilon\}$, with t as the parameter, satisfies condition (L$'$) at $t = 0$, and has Fisher information $i_{\phi_0}(\mathbf{v}) > 0$.

Let $\lambda : \Phi \to R$ be a scalar function on Φ. For any $\mathbf{v} \in V_0$, we say that λ is pathwise differentiable in direction $\mathbf{v}$ if $\lambda(t) = \lambda(\phi_0 + t\mathbf{v})$ is continuously differentiable in t for $|t| < \epsilon$. Let $V_1 \subset V_0$ be the set of $\mathbf{v} \in V_0$ such that $\lambda(\cdot)$ is pathwise differentiable in direction $\mathbf{v}$ and $\lambda_0'[\mathbf{v}] = \frac{d}{dt}\lambda(t)|_{t=0} \neq 0$.

Let T_n be an estimator of $\lambda(\phi)$. For any $\mathbf{v} \in V_1$, we say that T_n is pathwise regular in direction $\mathbf{v}$ if for any $\rho > 0$ and $t_n \to 1$ we have

$$\limsup P_{\phi_{n,\rho}}(T_n < \lambda(\phi_{n,\rho})) \leq \liminf P_{\phi_{n,-\rho}}(T_n < \lambda(\phi_{n,-\rho}))$$

where

$$\phi_{n,\rho} = \phi_0 + \frac{t_n\rho}{\sqrt{n}}\mathbf{v}, \quad \phi_{n,-\rho} = \phi_0 - \frac{t_n\rho}{\sqrt{n}}\mathbf{v}.$$

Let $V_2 \subset V_1$ be the set of directions in which T_n is pathwise regular.

Proposition 15. *Let T_n and V_2 be defined as above and assume that V_2 is nonempty. Then, for any $\rho > 0$*

$$\limsup P_{\phi_0}(\sqrt{n}|T_n - \lambda(\phi_0)| < \rho) \leq P(|Z_0| < \rho)$$

where Z_0 is a normal variable with mean zero and variance $\sigma^2 = \sup_{\mathbf{v}\in V_2} \frac{|\lambda_0'[\mathbf{v}]|^2}{i_{\phi_0}(\mathbf{v})}$.

Proposition 15 follows from the application of Proposition 14 to each family $\{P_{\phi_0+t\mathbf{v}}; |t| < \epsilon(v)\}$ induced by each $\mathbf{v} \in V_2$. The idea that a bound for the performance of estimators of a scalar function of an infinite dimensional parameter can be obtained by the application of the one dimensional theory to each one dimensional subfamily is due to Stein (1956).

Very often, the family $\{P_\phi\}$ is sufficiently smooth so that the space V_0 (with the addition of the zero vector) defined above is a linear space, and there is an inner product $\langle \mathbf{v}_1, \mathbf{v}_2\rangle$ on V_0 such that $i_{\phi_0}(\mathbf{v}) = \langle \mathbf{v}, \mathbf{v}\rangle = \|\mathbf{v}\|^2$ (see Wong and Severini (1991)). The quantities $\langle \mathbf{v}_1, \mathbf{v}_2\rangle$ and $\|\mathbf{v}\|$ are called the Fisher information inner product (of $\mathbf{v}_1$ and $\mathbf{v}_2$) and the Fisher information norm (of $\mathbf{v}$) respectively. If $\lambda(\phi)$, as a function on the normed linear space $(V_0, \|\cdot\|)$, is differentiable in a strong enough sense so that $\lambda(\cdot)$ is pathwise regular in each direction $\mathbf{v} \in V_0$, then the variance σ^2 in Proposition 15 has a nice geometric interpretation as the dual norm of $\lambda_0'[\cdot]$ as an element in the dual space $(V_0', \|\cdot\|^*)$. This generalizes the familiar geometry of the finite dimensional case: when $\phi \in$

$\mathbf{R}^p$, we have $\|\mathbf{v}\|^2 = \mathbf{v}'I\mathbf{v}$ for $\mathbf{v} \in \mathbf{R}^p$ where I is the $p \times p$ Fisher information matrix; furthermore, $(\|\mathbf{u}\|^*)^2 = \mathbf{u}'I^{-1}\mathbf{u}$ for $\mathbf{u} \in \mathbf{R}^p$.

The problem of finding estimators which can attain such Fisher information bounds under reasonably general conditions is not entirely resolved. In the case when the parameter space is compact under an appropriate norm Wong and Severini (1991) provide conditions under which the bound is attained by the plug-in estimator $\lambda(\hat{\phi}_n)$, where $\hat{\phi}_n$ is an approximate maximum likelihood estimator of ϕ. Other constructions in the related problem of semiparametric estimation are discussed by Bickel et al. (1991).

7. Proofs

Proposition 1. This follows from Proposition 6.

Proposition 2. The proof of this proposition is included in that of Proposition 4.

Proposition 3. There are already several proofs in the literature. The shortest seems to be the one based on an idea of Bickel's, as reported in Roussas (1972). We outline an alternative proof because it seems to offer some additional insights (see the remark before Proposition 5). Let

$$Y_n = \sqrt{n}(T_n - \theta_0), \quad \Delta_n = \Delta_{n,\theta_0}.$$

Then, under condition (L), $\{\mathcal{L}(Y_n, \Delta_n|\theta_0),\ n = 1, 2, \ldots\}$ is a tight sequence. (Here $\mathcal{L}(W|\theta)$ denotes the law of a variable W when θ is the parameter value.) By a contiguity argument, for any $h \in R$, we can choose a subsequence $\{n\}'$ such that along this subsequence,

$$\mathcal{L}(Y_n, \Delta_n|\theta_{n,h}) \to \mathcal{L}_h(Y, \Delta),$$

where the limiting law can depend on h. We now need a technical lemma.

Lemma. *Let (Y, Z) be two random variables whose joint distribution depends on a real parameter h. Suppose*

a) $\mathcal{L}_h(Z) = N(h, C)$, *where C is a constant,*
b) $\mathcal{L}_h(Y|Z)$ *does not depend on h,*
c) $\mathcal{L}_h(Y - h)$ *does not depend on h.*

Then $Y - Z$ is independent of Z, for all h.

Proof of Lemma. For simplicity, let $C = 1$ in (a).

By (a) and (c),

$$D_1 = E_h(e^{i(Z-h)}) \quad \text{and} \quad D_2 = E_h(e^{i(Y-h)})$$

are constants independent of h.

By (b), $g(Z) = E_h(e^{i(Y-Z)}|Z)$ does not depend on h. We need to show that, furthermore, $g(Z)$ does not depend on Z. Let $f(Z) = (g(Z) - D_2/D_1)$, then

$$\begin{aligned} E_h(f(Z)e^{i(Z-h)}) &= E_h(g(Z)e^{i(Z-h)}) - D_2 \\ &= E_h(e^{i(Y-h)}) - D_2 = 0. \end{aligned}$$

Hence $\int f(z)e^{i(z-h)}\phi(z-h)dz = 0$ for all h, where $\phi(z)$ is the standard normal density. Thus, the convolution of $f(z)$, with the function $\theta(z) = e^{iz}\phi(z)$, is identically zero. Since $\theta(z)$ has a Fourier transform which is nonzero everywhere, we must have $f(z) = 0$ for almost all z, i.e., $g(Z) = D_2/D_1$ as required.

To complete the proof of Proposition 3, let $Z = \Delta/i_\theta$; then it suffices to check that the conditions of the lemma are satisfied: Condition (a) follows because, by contiguity arguments,

$$\mathcal{L}(\Delta_n|\theta_{n,h}) \to N(i_\theta h, i_\theta); \text{ hence } \mathcal{L}_h(Z) = N(h, i_\theta^{-1}).$$

Condition (c) follows from the assumption that $\mathcal{L}\big(\sqrt{n}(T_n - \theta_{n,h})|\theta_{n,h}\big)$ converges to a limit law independent of h. Finally, condition (b) can be established by the following argument. Let

$$Z_n = (\Delta_n h + R_{n,h})/i_\theta h, \quad Z = \Delta/i_\theta.$$

Then

$$\mathcal{L}(Y_n, Z_n|\theta_{n,h}) \to \mathcal{L}_h(Y, Z).$$

For any bounded continuous $f(\cdot)$ and $g(\cdot)$,

$$\begin{aligned} &E(f(Y_n)g(Z_n)) \\ &= \int f(y_n)g(z_n)dP_{\theta_{n,h}}(\mathbf{X}_n) \\ &= \int f(y_n)g(z_n)e^{z_n i_\theta h + \frac{1}{2} i_\theta h^2} dP_{\theta_0}(\mathbf{X}_n) \\ &\to \int f(y)g(z)e^{z i_\theta h + \frac{1}{2} i_\theta h^2} dP_0(Y, Z) \\ &= \int f_1(z)g(z)e^{z i_\theta h + \frac{1}{2} i_\theta h^2} dP_0(Z) = E_h(f_1(Z)g(Z)) \end{aligned}$$

where

$$f_1(Z) = E_0(f(Y)|Z).$$

On the other hand, $E(f(Y_n)g(Z_n)) \to E_h(E_h(f(Y)|Z)g(Z))$. Hence

$$E_h(E_h(f(Y)|Z)g(Z)) = E_h(f_1(Z)g(Z))$$

for all continuous $g(\cdot)$, $f(\cdot)$ and h, and condition (b) follows.

Proposition 4. Following the approach of Bahadur (1964), we deduce Fisher's bound from the optimality of the likelihood ratio (LR) test. We use the following standard consequences of condition (L):

i) $\frac{\Delta_{n,\theta}}{i_\theta} \xrightarrow{\mathcal{D}} N(0, i_\theta^{-1})$ under P_θ.

ii) For any $h \in \mathbf{R}$, $\frac{\Delta_{n,\theta}}{i_\theta} \xrightarrow{\mathcal{D}} N(h, i_\theta^{-1})$ under $P_{\theta_{n,h}}$.

For any fixed $\rho > 0$ let

$$a = \limsup P_{\theta_{n,\rho}}(T_n < \theta_{n,\rho})$$

and

$$b = \liminf P_{\theta_{n,-\rho}}(T_n < \theta_{n,-\rho}).$$

We will prove the desired inequality for rational $\rho > 0$, which will then imply the result for all real ρ by monotonicity and continuity considerations.

a) First, consider testing

$$H_0 : P_{\theta_{n,\rho}} \quad \text{versus the alternative} \quad H_A : P_{\theta_0}$$

based on the data $\mathbf{X}_n = (X_1, \ldots, X_n)$. Let $B_{n,\rho} = \{T_n < \theta_{n,\rho}\}$ be a rejection region. Then, by assumption, $\limsup\{\text{size}(B_{n,\rho})\} = a$. Let $A_{n,\rho} = \{\Delta_n \rho + R_n < i\rho^2 + \sqrt{i\rho^2} z_{\alpha_n}\}$ where Δ_n, i, R_n are as in condition (L) with $\theta = \theta_0$, and z_{α_n} is the lower α_n quantile of the standard normal distribution. Let $\alpha_n \downarrow a' > a$; then it follows from the same arguments that establish (i) and (ii) that

$$\text{size}(A_{n,\rho}) \to a' > a = \limsup \text{size}(B_{n,\rho}).$$

Since $A_{n,\rho}$ is a LR rejection region, we must have

$$\begin{aligned}
\limsup & P_{\theta_0}(\sqrt{n}(T_n - \theta_0) < \rho) \\
&= \limsup\{\text{power}(B_{n,\rho})\} \\
&\le \lim P_{\theta_0}(A_{n,\rho}) = P\big(Z_{\theta_0} < \rho + \frac{z_{a'}}{\sqrt{i}}\big).
\end{aligned}$$

b) Next, consider testing

$$H_0 : P_{\theta_0} \quad \text{versus the alternative} \quad H_A : P_{\theta_{n,-\rho}}.$$

Let $B_{n,-\rho} = \{T_n < \theta_{n,-\rho}\}$; then, by assumption,

$$\liminf\{\text{power}(B_{n,-\rho})\} = b.$$

Let $A_{n,-\rho} = \{\Delta_n(-\rho) + R_n > i\rho^2 + \sqrt{i\rho^2} z_{\alpha_n}\}$, and let $\alpha_n \to 1 - b'$ where $b' < b$. Then,

$$\text{power}(A_{n,-\rho}) = P_{\theta_{n,-\rho}}(A_{n,-\rho}) \to b' < b.$$

Since $A_{n,-\rho}$ is a LR rejection region, we have

$$\begin{aligned}\liminf P_{\theta_0}(\sqrt{n}(T_n - \theta_0) < -\rho) &= \liminf P_{\theta_0}(B_{n,-\rho}) \geq \lim P_{\theta_0}(A_{n,-\rho}) \\ &= P\Big(Z_{\theta_0} < -\rho + \frac{z_{b'}}{\sqrt{i}}\Big).\end{aligned}$$

c) Since the above inequalities hold for any $a' > a$ and $b' < b$, we have

$$\begin{aligned}\limsup P_{\theta_0}(\sqrt{n}|T_n - \theta_0| < \rho) &\leq P\big(-\rho + \frac{z_b}{\sqrt{i}} < Z_{\theta_0} < \rho + \frac{z_a}{\sqrt{i}}\big) \\ &\leq P(|Z_{\theta_0}| < \rho).\end{aligned}$$

The last inequality follows because $z_a \leq z_b$ and the normal distribution has higher concentrations for symmetric intervals than asymmetric intervals of the same length.

Special but straightforward arguments are needed for the case when a or b take the values 0 or 1. Finally, if $a = b = \frac{1}{2}$ then $z_a = z_b = 0$ and we may test

$$H_0 : P_{\theta_0} \quad \text{versus} \quad H_A : P_{\theta_{n,-\rho_1}}$$

in part (b), where ρ_1 need not be the same as ρ in part (a). This will give, by exactly the same arguments as above, Fisher's bound for concentration probability for all asymmetric intervals.

Proposition 5. Replace the technical lemma in the proof of Proposition 3 by the following easy lemma.

Lemma. *Suppose (Y, Z) have a joint distribution depending on h, and*
a) $\mathcal{L}_h(Z) = N(h, C)$, *where C is a constant,*
b) $\mathcal{L}_h(Y|Z)$ *does not depend on h,*
c) $\mathcal{L}_h(Y - h)$ *has mean value independent of h.*
Then $\text{Var}(Y) \geq \text{Var}(Z)$ *under $\mathcal{L}_h(Y, Z)$, for all values of h.*

Proof of Lemma. Let $E_h(Y - h) = D$; then D is a constant by condition (c). Hence $Y - D$ is an unbiased estimate for h. The result then follows from the Rao-Blackwell argument.

Proposition 6. Let $\gamma(\theta) = \lim P_\theta(T_n < \theta)$ and

$$f_n(\theta) = \begin{cases} P_\theta(T_n < \theta) - \gamma(\theta) & \text{if } \theta \in \Omega \\ 0 & \text{otherwise,} \end{cases}$$

$f_n(\cdot)$ is measurable since, under condition (M), $P_\theta(T_n < \theta)$ is measurable in θ. Then, for any rational h,

$$\int f_n(\theta + h/\sqrt{n}) \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta^2} d\theta$$
$$= \int f_n(t) e^{ht/\sqrt{n} - h^2/2n} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt.$$

This last term goes to zero by dominated convergence because $f_n(\theta) \to 0$ almost everywhere. Thus, if we define $f_{n,h}(\theta) = f_n(\theta + h/\sqrt{n})$, then $f_{n,h}(\cdot) \to 0$ in standard Gaussian measure, and hence there exist a subsequence along which $f_{n,h}(\cdot) \to 0$ a.e. By diagonal extraction, we arrive at a subsequence along which

$$f_{n,h}(\cdot) \xrightarrow{\text{a.e.}} 0 \quad \text{for all rational } h.$$

i.e.

$$P_{\theta+h/\sqrt{n}}(T_n < \theta + h/\sqrt{n}) - \gamma(\theta + h/\sqrt{n}) \to 0 \text{ a.e.} \quad \text{for all rational } h.$$

Since $\gamma(\cdot)$ is continuous almost everywhere, we see that, for this subsequence, the condition of Proposition 4 is satisfied for almost all θ. The result follows.

Proposition 7. Since $P(|Z_\theta| < \rho)$ is continuous in ρ, it suffices to prove the result for rational ρ. Thus, let ρ be a fixed rational number, and define

$$g_n(\theta) = P_\theta(\sqrt{n}|T_n - \theta| < \rho);$$
$$g(\theta) = P(|Z_\theta| < \rho),$$
$$g_\infty(\theta) = \lim g_n(\theta).$$

Under condition (M), all these are measurable functions. We wish to prove that $g_\infty(\cdot) \leq g(\cdot)$ a.e. By Lemma (C) in the proof of Proposition 12 and dominated convergence, we have

$$\int_I (g_\infty(\theta) - g(\theta)) d\theta \leq 0 \quad \text{for any interval } I.$$

For any $\delta > 0$, let

$$\mathcal{D}_\delta = \{\theta \in \Omega : g_\infty(\theta) - g(\theta) > \delta\}.$$

For any $\epsilon > 0$, we can find U = a union of finite open intervals such that

$$\mu(\mathcal{D}_\delta \Delta \mathsf{U}) < \epsilon.$$

Then

$$\delta\mu(\mathcal{D}_\delta) < \int_{\mathcal{D}_\delta} (g_\infty - g) d\theta \leq \int_{\mathsf{U}} (g_\infty - g) d\theta + 2\epsilon$$
$$\leq 2\epsilon,$$

which implies $\mu(\mathcal{D}_\delta) = 0$. Since $\delta > 0$ is arbitrary, we have $\mu\{\theta \in \Omega : g_\infty(\theta) \leq g(\theta)\} = 1$.

Proposition 8. This follows from Proposition 7 and Egorov's theorem.

Propositions 9 and 10. These two propositions are equivalent, and Proposition 10 follows trivially from Proposition 11. To see the equivalence, suppose Proposition 10 is true. Then, for any fixed $\epsilon > 0$ and $\delta > 0$, we can find $n_0 > 0$ and

$$\theta_n \in I_\delta \cap A_n(\epsilon) \quad \text{for all } n \geq n_0.$$

(Here, $A_n(\epsilon) = \{\theta : g_n(\theta) \leq g(\theta) + \epsilon\}$ and $I_\delta = \{\theta : |\theta - \theta_0| < \delta\}$). Hence,

$$\limsup_{n\to\infty} \inf_{\theta\in I_\delta} g_n(\theta) < \sup_{\theta\in I_\delta} g(\theta) + \epsilon.$$

Proposition 9 follows since $g(\cdot)$ is continuous in θ. Conversely, suppose the conclusion of Proposition 10 is not true. Then, there is an $\epsilon > 0$, an interval I and a subsequence $\{n\}' \subset \{n\}$, such that

$$A_n(\epsilon) \cap I = \phi \quad \forall\, n \in \{n\}'.$$

Pick $\theta_0 \in$ interior of I, and δ so small that $I_\delta = \{|\theta - \theta_0| < \delta\} \subset I$, then

$$g_n(\theta) > g(\theta) + \epsilon \quad \forall\, \theta \in I_\delta,\ n \in \{n\}'.$$

Hence,

$$\limsup_{n\to\infty} \inf_{\theta\in I_\delta} g_n(\theta) \geq \inf_{\theta\in I_\delta} g(\theta) + \epsilon,$$

or

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \Big(\inf_{\theta\in I_\delta} g_n(\theta)\Big) \geq g(\theta_0) + \epsilon,$$

and Proposition 9 is untrue.

Proposition 11. This follows from Proposition 12.

Proposition 12. Define

$$a_n(s; \mathbf{X}_n) = P\Big(\sqrt{n}(\Theta - \hat{\theta}_n) < s \;|\mathbf{X}_n, \pi\Big)$$

and

$$k_n(\mathbf{X}_n) = P(\sqrt{n}|\Theta - T_n| < \rho \;|\mathbf{X}_n, \pi).$$

The proof can be completed in three steps.

Lemma (A). *Under condition* (P), *let* $\theta \in I_\pi$ = *interior of* $\{\theta : \pi(\theta) > 0\}, F(s) = P(Z_\theta < s)$. *Then,*

$$\sup_s |a_n(s; \mathbf{X}_n) - F(s)| \xrightarrow{P} 0 \text{ under } P_\theta(\mathbf{X}_n).$$

Proof. Let $\epsilon > 0$, $\delta > 0$ be arbitrary. Find K so large that $F(K) > 1 - \epsilon$, $F(-K) < \epsilon$. By monotonicity of a_n and F in s, it follows from condition (P) that there is an $m_0 > 0$ such that

$$P_\theta\Big(\sup_{|s|>K} |a_n(s; \mathbf{X}_n) - F(s)| < 2\epsilon\Big) > 1 - \delta/2 \quad \forall\, n > m_0.$$

Since $F(\cdot)$ is continuous, we can find a $\Delta > 0$ such that

$$\sup_{\substack{s_1, s_2 \in [-K, K] \\ |s_2 - s_1| < \Delta}} |F(s_2) - F(s_1)| < \epsilon.$$

Now divide $[-K, K]$ into M intervals each of length less than Δ, and let $\rho_0 < \rho_1 < \cdots < \rho_M$ be the dividing points. Then there exists $m_1 > 0$ such that

$$P_\theta\Big(|a_n(\rho_i; \mathbf{X}_n) - F(\rho_i)| < \epsilon,\ \forall\, i = 0, \ldots, M\Big) > 1 - \delta/2 \quad \forall\, n > m_1.$$

It then follows from monotonocity of $a_n(\cdot; X_n)$ and $F(\cdot)$ that

$$P_\theta\Big(\sup_{|s| \le K} |a_n(s; \mathbf{X}_n) - F(s)| < 3\epsilon\Big) > 1 - \delta/2,\ \forall\, n > m_1.$$

Hence, $n > \max(m_0, m_1) \Rightarrow P_\theta\Big(\sup_s |a_n(s; \mathbf{X}_n) - F(s)| < 3\epsilon\Big) > 1 - \delta$, which proves Lemma (A).

Lemma (B). *Under condition* (P), *let* $\theta \in I_\pi$; *then for any* $\epsilon > 0$, $\delta > 0$, *there exist a* m_0 *such that*

$$P_\theta\Big(k_n(\mathbf{X}_n) \le g(\theta) + \epsilon\Big) > 1 - \delta \quad \text{for all } n > m_0.$$

Proof.

$$\begin{aligned}
k_n(\mathbf{X}_n) &= P(\sqrt{n}|\Theta - T_n| < \rho | \mathbf{X}_n, \pi) \\
&= P(T_n - \rho/\sqrt{n} < \Theta < T_n + \rho/\sqrt{n} | \mathbf{X}_n, \pi) \\
&= P\Big(\sqrt{n}(T_n - \hat{\theta}_n) - \rho < \sqrt{n}(\Theta - \hat{\theta}_n) < \sqrt{n}(T_n - \hat{\theta}_n) + \rho | \mathbf{X}_n, \pi\Big) \\
&\le \sup_s [a_n(s + \rho; \mathbf{X}_n) - a_n(s - \rho; \mathbf{X}_n)] \\
&\le \sup_s [F(s + \rho) - F(s - \rho)] + 2 \sup_s |a_n(s; \mathbf{X}_n) - F(s)|.
\end{aligned}$$

The first term is bounded by $g(\theta)$ because of the shape of the normal density. The lemma follows by applying Lemma (A) to bound the second term.

Lemma (C). *Suppose condition* (P) *holds. Then,* $\limsup \int \pi(\theta)\big(g_n(\theta) - g(\theta)\big)d\theta \leq 0$ *for any piecewise continuous prior density* $\pi(\cdot)$.

Proof.

$$\begin{aligned}\int \pi(\theta)g_n(\theta)d\theta &= \int \pi(\theta)P(\sqrt{n}|T_n - \theta| < \rho|\Theta = \theta)d\theta \\ &= \int P\Big(\sqrt{n}|\Theta - T_n(\mathbf{x}_n)| < \rho|\mathbf{X}_n = \mathbf{x}_n, \pi\Big)p_\pi(\mathbf{x}_n)d\nu_n(\mathbf{x}_n) \\ &= \int k_n(\mathbf{x}_n)p_\pi(\mathbf{x}_n)d\nu_n(\mathbf{x}_n).\end{aligned}$$

Here $p_\pi(\mathbf{x}_n) = \int P_{\theta'}(\mathbf{x}_n)\pi(\theta')d\theta'$ is the marginal density of $\mathbf{X}_n$ under prior $\pi(\cdot)$. Let

$$K_n(\theta') = E_{\theta'}\Big(k_n(\mathbf{X}_n)\Big) = \int k_n(\mathbf{x}_n)p_{\theta'}(\mathbf{x}_n)d\nu_n(\mathbf{x}_n),$$

then

$$\int \pi(\theta)g_n(\theta)d\theta = \int \pi(\theta)K_n(\theta)d\theta.$$

For any $\theta \in I_\pi$, and $\epsilon > 0, \delta > 0$ arbitrary, it follows from Lemma (B) that there exists m_0 such that $n > m_0$ implies

$$\begin{aligned}K_n(\theta) &\leq \int_{\{k_n(\mathbf{X}_n) \leq g(\theta)+\epsilon\}} k_n(\mathbf{x}_n)dP_\theta(\mathbf{x}_n) \\ &\quad + P_\theta\Big(k_n(\mathbf{X}_n) > g(\theta) + \epsilon\Big) \\ &\leq g(\theta) + \epsilon + \delta,\end{aligned}$$

and hence,

$$\limsup K_n(\theta) \leq g(\theta).$$

Thus, by Fatou's lemma, $\limsup \int \pi(\theta)g_n(\theta)d\theta \leq \int \pi(\theta)g(\theta)d\theta$. We have, in fact, proven a stronger statement than Proposition 12.

Proof of Proposition 14. By the inverse function theorem, locally θ and λ are one to one differentiable functions of each other. Reparameterize by λ and write $P_\lambda = P_{\theta(\lambda)}, \lambda_0 = \lambda(\theta_0)$. Let $\lambda_n = \lambda_0 + h/\sqrt{n}$. Then,

$$\theta_n = \theta(\lambda_n) = \theta_0 + \frac{d\theta}{d\lambda}(\tilde{\lambda}_n) \cdot \frac{h}{\sqrt{n}}$$

where $\tilde{\lambda}_n$ is between λ_0 and λ_n, and

$$\frac{d\theta}{d\lambda}(\tilde{\lambda}_n) \to \frac{d\theta}{d\lambda}(\lambda_0) = \lambda'(\theta_0)^{-1}.$$

By condition (L′),

$$\frac{dP_{\lambda_n}}{dP_{\lambda_0}} = \frac{dP_{\theta(\lambda_n)}}{dP_{\theta_0}} = e^{\lambda'(\theta_0)^{-1}h\Delta_n - \frac{1}{2}\lambda'(\theta_0)^{-2}i_{\theta_0}h^2 + R_n}.$$

Hence, condition (L) is satisfied for the family $\{P_\lambda\}$ with λ as the parameter and $\lambda'(\theta_0)^{-2}i_{\theta_0}$ as the Fisher information. Proposition 14 now follows from Proposition 4.

Acknowledgement

The author is grateful to Professors R. R. Bahadur and Charles Stein for very helpful conversations, and to Professor L. LeCam for his written reply to a number of queries. This research was supported in part by a Guggenheim Fellowship and in part by National Science Foundation Grant No. DMS 86-01732. This manuscript was prepared using computer facilities supported in part by the National Science Foundation Grants DMS 89-05292, DMS 87-03942, DMS 86-01732, and DMS 84-04941 awarded to the Department of Statistics at The University of Chicago, and by The University of Chicago Block Fund.

References

Bahadur, R. R. (1964). On Fisher's bound for asymptotic variances. *Ann. Math. Statist.* **35**, 1545-1552.

Basu, D. (1956). The concept of asymptotic efficiency. *Sankhyā* **17**, 193-196.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1991). Efficient and adaptive inference in semiparametric models. Johns Hopkins University Press.

Edgeworth, F. Y. (1908-9). On the probable errors of frequency-constants. *J. Roy. Statist. Soc.* **71**, 381-397, 499-512, 651-678. Addendum *ibid.* **72**, 81-90.

Fisher, R. A. (1925). Theory of statistical estimation. *Proc. Cambridge Philos. Soc.* **22**, 700-725.

Hajek, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. verw. Gebiete* **14**, 323-330.

Hajek, J. (1972). Locally asymptotic minimax and admissibility in estimation. *Proc. 6th Berkeley Symp. Math. Statist. Probab.* **1**, 175-194.

LeCam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. California Publ. Statist.* **1**, 277-330.

LeCam, L. (1958). Les propriets asymptotiques des solutions de Bayes. *Publ. Inst. Statist. Univ. Paris* **7**, 17-35.

LeCam, L. (1960). Locally asymptotically normal families of distributions. *Univ. California Publ. Statist.* **3**, 27-98.

Pratt, J. W. (1976). F. Y. Edgeworth and R. A. Fisher on the efficiency of maximum likelihood estimation. *Ann. Statist.* **4**, 501-514.

Roussas, G. G. (1972). *Contiguity of Probability Measures: Some Applications in Statistics.* Cambridge University Press, Cambridge.

Stein, C. (1956). Efficient nonparametric testing and estimation. *Proc. 3rd Berkeley Symp. Math. Statist. Probab.* **1**, 187-196, University California Press.

Wong, W. H. and Severini, T. A. (1991). On maximum likelihood estimation in infinite dimensional parameter spaces. *Ann. Statist.* **19**, 603-632.

Department of Statistics, University of Chicago, 5734 University Ave., Chicago, IL 60637, U.S.A.

(Received August 1989, accepted August 1991)